

Weekly Report

Lu Junhua

2016 年 12 月 5 日

Done

- Discuss more details about bot detection.

暂时研究集中在全程挂, 因为这是比较好检测的(全程都在用). 从这里入手比较好. 暂时不要求太高的时效性.

常用一些特征: 金钱档, 经验档, 创建档, 登陆档, 升级档, 任务领取完成.

登录登出不匹配问题, 类似于QQ异地登陆被踢.

作坊出来的一般是一起的, 属性会比较接近, 至少现在是这样的. 可以尝试使用聚类的方法来做这个东西.

提取特征时候, 记住不仅要用一个动作结果值, 还要考虑参数值.

最终目标实现完全自动化, 不再人工去找不断变化的特征.

- 对此一些个人看法:

这种外挂号生命周期是怎么样的? 能否作一些简单的时间线对齐? 正如之前做的虚拟生活中的时间经历一样. 如果时间周期不长那种, 对齐可以做一些可视化方面工作.

作为训练的输入, 数据的处理、清洗, 数据的监控都可以做为我们后面做的内容, 但是领域需求实在不算明确. 后面要跟进一步了解他们的需求, 有针对性的设计我们后面要做的内容.

如果后面用神经网络, 我们能否在这方面做点文章? (但这个好像又偏向深度学习领域了)

现在按照作坊的这种风格, 可以试着用那个张量方法了(不用张量分解那个), 探测dense block.

- Read materials on clustering, time-series mining, tutorial on anomaly detection, etc.
- Gained more experiences in Qiannv Youhun.

To do

- I may get interviewed and take on internship in Netease.

论文

- Ph.D. thesis *EventFlow: Visual Analysis of Temporal Event Sequences* 仍然是对于事件的对齐, 整合, 聚合, 但是加入了一些事件合并, 以不同方式对齐的等交互, 其功能和之前设计的一些Flow 确实有区别之处, 也可以用作异常情况检测一些方法. 在网易那边简单演示时, 他们也认为在现阶段用作监控是有一定的价值的.

- VAST2016/TVCG *A Visual Analytics Approach for Understanding Reasons behind Snowballing and Comeback in MOBA Games* moba类型游戏中一些行为的可视分析. 系统比较复杂, 编码信息非常多, 但是从单个视图来看比较易懂. 问题明确也可以看出来是与领域专家有密切合作的. 仍然比较羡慕他们对于glyph的使用, 作为一个全局小视图来用. case相对来说比较松散, 但是部件差不多都用上了.
- 下面是今年VIS 16 The event event workshop一些小文章, 有些还是有点启发的.

The Event Quartet: How Visual Analytics Works for Temporal Data: 提出对于常见需要可视化的Event分类: 单时间线多事件类型; 多时间线单事件; 多时间线多事件; 后两者仍然处于探索阶段.

Soft patterns: Moving Beyond Explicit Sequential Patterns During Visual Analysis of Longitudinal Event Datasets 很多现有方法太关注严格序列限制, 导致错过一些仍然被隐藏的很需要被理解的(在现实数据中). 现在一直有的Hard pattern: ‘Event A followed by Event B followed by Event C.’ 精确而严格; 然而, 有时候严格的限制会太限制性了. 对于小的事件的变化太敏感, 没有太大的domain meaning. 所以定义了soft patterns: groups of temporally proximate events which unfold(展现) over time, but potentially without the strict ordering and temporal constraints used in much of the prior visualization work for event sequence analysis. 作者认为这样的一种pattern如能更好的发掘并展示, 是十分有意思也有价值的.

ChronoView: A Method to Overview Many Temporal Patterns 这个节省空间的设计其实可以当作是某种glyph. 一个圆周是时间轴, 同一个时间组不同时间段发生有多个时间点, 取其重心作为一个编码的点, 可以对这个点进行加工修饰, 编码其他信息. 为了减少视觉混乱, 还可以采用一些影射函数达到那种弦图里面绑定的效果.

Sequence Pre-processing: Focusing Analysis of Log Event Data 本文倾向于描述时序数据的预处理(诸如取出噪声等), 其实其方法有些在Fan Du今年那篇综述中也有提到. 移除时间、设置代理、选择合适session、session细分这四种方法. 这些方法需要人知识的结合. 再回到我们今年想做的东西上来, 这几个方法可以作为初始入手或者说提供一些交互上设计.